

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2003, 19:44:49 ; Search time 16.0286 seconds  
(without alignments)  
2650,980 Million cell updates/sec

Title: US-09-497-967-6

Perfect score: 2342

Sequence: 1 MKYNILLIILISLFINELRA.....STTFAKFLISLILFISFYLL 442

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : PIR 73:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2083	88.9	395	2	A46031	Immobilization sur
2	213.5	9.1	713	2	A35502	major surface-labe
3	212	9.1	1274	2	T42017	cysteine rich prot
4	210.5	9.0	557	2	A48434	variant-specific s
5	210.5	9.0	596	2	A45664	variant-specific s
6	206	8.8	1766	2	A42125	trophozoite cystei
7	205	8.8	1459	2	T32271	hypothetical prote
8	203	8.7	677	2	C42125	trophozoite cystei
9	202.5	8.6	3635	2	T10053	laminin alpha-5 ch
10	198	8.5	3084	1	MMMSA	laminin alpha-1 ch
11	195.5	8.3	1372	2	T25933	hypothetical prote
12	189	8.1	5376	2	T42215	zonadhesin - mouse
13	183.5	7.8	1895	2	T15881	hypothetical prote
14	183	7.8	1299	2	T43251	furin (EC 3.4.21.7
15	182	7.8	667	2	A48579	trophozoite surfac
16	181.5	7.7	3075	2	S14458	laminin alpha-1 ch
17	180.5	7.7	2395	1	S50820	surface protein ty
18	178	7.6	3106	1	S53868	laminin alpha-2 ch
19	177.5	7.6	1607	1	MMMS82	laminin gamma-1 ch
20	177	7.6	439	2	A36385	surface antigen se
21	173	7.4	1790	1	MMFEB1	laminin beta-1 cha
22	172	7.3	899	2	G02428	subtilisin-like pr
23	172	7.3	915	2	JC6148	subtilisin-like pr
24	170.5	7.3	2195	2	T34264	hypothetical prote
25	169.5	7.2	1169	2	S38181	floculation prote
26	169	7.2	1548	2	S34583	serine proteinase
27	166.5	7.1	1557	2	T28811	hypothetical prote
28	166.5	7.1	1797	2	T21889	hypothetical prote
29	166.5	7.1	1805	2	T21888	hypothetical prote

30	165.5	7.1	507	2	T44768	antifreeze glycope
31	165.5	7.1	739	2	B88553	protein KOAH4.2b [
32	165.5	7.1	1680	2	A43434	furin (EC 3.4.21.7
33	165.5	7.1	1700	2	S08167	Balbani ring 3 pr
34	165.5	7.1	3712	2	S18253	laminin alpha-1 ch
35	164	7.0	1252	2	S36016	oocyst wall protei
36	164	7.0	4776	2	E95206	cell wall surface
37	162.5	6.9	2704	2	S09118	G surface protein
38	160	6.8	1609	1	MMHUB2	laminin gamma-1 ch
39	158.5	6.8	600	2	S07638	spore coat protein
40	158	6.7	738	2	S40992	hypothetical prote
41	157	6.7	822	2	A38420	antifreeze glycopr
42	156.5	6.7	2823	2	T23064	hypothetical prote
43	156.5	6.7	2823	2	F87908	protein T22A3.8 [i
44	156.5	6.7	3102	2	T43291	laminin alpha chai
45	156	6.7	1192	2	S69000	laminin gamma 2 ch

ALIGNMENTS

RESULT 1

A46031

Immobilization surface I-antigen precursor - Ichthyophthirius multifiliis (fragment)  
C:Species: Ichthyophthirius multifiliis  
C>Date: 29-Sep-1999 #sequence\_revision 29-Sep-1999 #text\_change 07-Dec-1999  
C:Accession: A46031

R:Clark, T.G.; McGraw, R.A.; Dickerson, H.W.  
Proc. Natl. Acad. Sci. U.S.A. 89, 6363-6367, 1992  
A:Title: Developmental expression of surface antigen genes in the parasitic ciliate I  
A:Reference number: A46031; MUID:92335298; PMID:1631132  
A:Accession: A46031  
A:Molecule type: mRNA; protein  
A:Residues: 1-395 <CLA>  
A:Cross-references: GB:M92907; NID:g3628568; PIDN:AAC36158.1; PID:g3628569  
A>Note: the authors translated the codon UUG for residue 330 as Ile  
A>Note: sequence extracted from NCBI backbone (NCBIN:108734, NCBIPI:108735); the sequ  
C:Genetics:  
A:Genetic code: SGC5  
C:Keywords: glycoprotein; surface antigen  
F:2-395/Product: Immobilization surface I-antigen #status experimental <MAT>  
F:156,191,245,281/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 88.9%; Score 2083; DB 2; Length 395;  
Best Local Similarity 99.2%; Pred. No. 4.9e-125;  
Matches 387; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY	20	AVPCPDGTQTQAGLTDVGAADLGTVCNRPNFYNGGAQGEANGNPFANNAARGICV	79
DB	1	AVPCPDGTQTQAGLTDVGAADLGTVCNRPNFYNGGAQGEANGNPFANNAARGICV	60
QY	80	PCQINRVGSVTNAGDLATLATQCSQTCTGTALDDGVTDFDRAAOCVKCKPNFYNGG	139
DB	61	PCQINRVGSVTNAGDLATLATQCSQTCTGTALDDGVTDFDRAAOCVKCKPNFYNGG	120
QY	140	SPQGEAPGVQVFAAGAAAGVAANTVSCVPCQLNKNDSPATAGAAANLATQCSNQCPGT	199
DB	121	SPQGEAPGVQVFAAGAAAGVAANTVSCVPCQLNKNDSPATAGAAANLATQCSNQCPGT	180
QY	200	VLDGVTLVFNTSATLCVKCRPNFYNGGSPQGEAPGVQVFAAGAAAGVAANTVSCVPC	259
DB	181	VLDGVTLVFNTSATLCVKCRPNFYNGGSPQGEAPGVQVFAAGAAAGVAANTVSCVPC	240
QY	260	QINKNDSPATAGAAANLATQCSQTCTGTATQDGVTLVFNSNSTQCSQCIANTFFNGNFE	319
DB	241	QINKNDSPATAGAAANLATQCSQTCTGTATQDGVTLVFNSNSTQCSQCIANTFFNGNLE	300
QY	320	AGKSQCLKCPVSKTTPAHAPGNATATQCLTTCPCAGTVLDDGVTSTNFVASATECTKCSA	379
DB	301	AGKSQCLKCPVSKTTPAHAPGNATATQCLTTCPCAGTVLDDGVTSTNFVASATECTKCSA	360
QY	380	GFFASKTTGTTAGTDTCTCKTLTGATA	409

Thu Feb 20 11:30:46 2003

A:Title: A Giardia duodenalis gene encoding a protein with multiple repeats of a toxin

A:Reference number: 222027  
A:Accession: T42017  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1274 <CHE>  
A:Cross-references: EMBL:L29079; NID:g951190; PID:g951191; PIDN:AAA74587.1  
A:Experimental source: specific host: Homo sapiens  
C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homol

Query Match 9.1%; Score 212; DB 2; Length 1274;  
Best Local Similarity 20.9%; Pred. No. 4.4e-06;  
Matches 97; Conservative 42; Mismatches 169; Indels 156; Gaps 21;

QY 23 CPDGTQAGLTDVCAADLGTVCNCRPNFYNGGAAGEANGNPF---AANNAARGICV 79  
DB 851 CVEGNAQQ-----CKTCRPGYINTDTKCTKDPAPCNVEGCETCVGNAQ 897  
QY 80 PCQINRVSVTNAGDLATLQCSCTQCTPGTALDDGVDFDRSAAOCVKCKPNFYNGG 139  
DB 898 QCKTCRPGYIN-----TDTKQC-TKDPEAPCNVEGCETCVGNAQCKTCRPGYINTD 951  
QY 140 SPQ-----GEAPGVQVFAAGAAAGVAATVSOCPQLNKNDSPATAGAAQLATQC---- 191  
DB 952 TKQCTKDPAP-----CNTPNCKTCDNFKTDN-----EICTKNDGD 988  
QY 192 ----SNOC-PTGTVLDDGVTLVFNSTATLCKVCRPNFYNGSGPGEAPGVQVFAAGAA 246  
DB 989 YLTPTNOCVPDCTAISG---YVGTDRKKACNPE-----CA 1022  
QY 247 AGVAATVSOCPQINLK-----NDSPATAGAAQLATQCS 281  
DB 1023 ECVGPANNOCTACPVGKMLQYTDTPVNGTCDMDOCSVSTNDGCAECGAOIGGTAYCS 1082  
QY 282 ----TQCTPT--GTATQDGVTLVFNSTQCSQCIANTF-----FNGFNFAKRSQCL-- 326  
DB 1083 KCKTOAPLNGCAASSRVAFCATITSGACTKCEGYFLKDGCGYQDTPGKOVCSNA 1142  
QY 327 ----KCPVSKTTPAHAPGNATATQCTCTCPAGTVLLDDGTSTNFVASATECTKCSAGF 381  
DB 1143 QGGNGKQCTCANGLAASDGNCA-----ECHSTCATCTAD-----AADKCKTCATGY 1189  
QY 382 FASKTTGFTAGTCTCTECTKLTSGATARYAPATQKVOCASST 425  
DB 1190 YKENGDDTTAG--LCKKCKSEKI-SGCK-----QCVSSS 1219

RESULT 4  
A48434  
variant-specific surface protein - Giardia lamblia (strain GS/M)  
C:Species: Giardia lamblia  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 07-May-1999  
C:Accession: A48434  
R:Nash, T.E.; Mowatt, M.R.  
Mol. Biochem. Parasitol. 51, 219-228, 1992  
A:Title: Characterization of a Giardia lamblia variant-specific surface protein (VSP  
A:Reference number: A48434; MUID:92244292; PMID:1574080  
A:Accession: A48434  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-557 <NAS>  
A:Cross-references: GB:M80480; NID:g159142; PID:g159143

Query Match 9.0%; Score 210.5; DB 2; Length 557;  
Best Local Similarity 22.9%; Pred. No. 2.7e-06;  
Matches 114; Conservative 48; Mismatches 184; Indels 151; Gaps 29;

QY 7 LILILSLFNELRVPCPDGTQQTQAGLTD-----VGAADLGTVCNCRPNFYNGGAQGE 61  
DB 1 MFLINCLITASTLAGAC---STTQANCVAEKCCEMVGTEI---CTQCKQNVIPINGVCEAA 55  
QY 62 ANGN---OPFAANNAARGICVPCQ-----INRVGSVTVNAGDLATL--ATQCSCTCPTGTAL 112

DB 361 GFFASKTTGFTAGTDTCTECTKLTSGATA 390

RESULT 2  
A35502  
major surface-labeled trophozoite antigen precursor - Giardia lamblia  
C:Species: Giardia lamblia  
C:Date: 09-Nov-1990 #sequence\_revision 09-Nov-1990 #text\_change 09-Sep-1997  
R:Gillin, F.D.; Hagblom, P.; Harwood, J.; Alley, S.B.; Reiner, D.S.; McCaffery, M.; So, M.  
Proc. Natl. Acad. Sci. U.S.A. 87, 4463-4467, 1990  
A:Title: Isolation and expression of the gene for a major surface protein of Giardia lam  
A:Reference number: A35502; MUID:90280395; PMID:2352929  
A:Accession: A35502  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-713 <GIL>  
A:Cross-references: GB:M33641; NID:g159131; PID:g159132  
C:Keywords: surface antigen; transmembrane protein

Query Match 9.1%; Score 213.5; DB 2; Length 713;  
Best Local Similarity 23.0%; Pred. No. 2.2e-06;  
Matches 127; Conservative 42; Mismatches 187; Indels 197; Gaps 31;

QY 42 GTCVNCNRPNFYNGGAQ-----GEANGNPPFAANNAARGI--CVPCQINRVGSVTN 91  
DB 111 GVCTEAPGVFAPVGAANTEQSVIAGDGTGVTIAAGNTYKGIADCAECSAPATGAE 170  
QY 92 AGDLATLATQC-----STQCTPTGT-----ALDD-----GV- 116  
DB 171 AGKAT--CTKCGVSKYLKDVCKRAQCNSTGNFVAVDDSENGKVCSCSDNLNGGVA 229  
QY 117 ---TDVDRSA--AQCCKCPNFY-----NGSPQGE-APGVQVFAAG 154  
DB 230 NCDTCSYDEQSKKIKTKTDNNLYKTSTSEGTSCVOKDQCKDGFPPKDDSSAGNKCLPCN 289  
QY 155 AAAAGVAATVSOCPQLNKNDSPA-----TAGAQLAT----OC 191  
DB 290 DSTDGLA-----NCATCALVSGRGAALVTCSTACTDGYKPSADKTTCEAVSNCKTPGCKAC 345  
QY 192 SNQCTPTGLVD--DGVTLVFNSTATL-----TQCTPTGTAI-QDGV-----TLVFSNS 301  
DB 346 SNEGKENEVCTDCDGTSTLPTSTQIDSCAKIGNYYGATEGAKKLCKECTAANCKTCDQD 405  
QY 216 ---CVKCRPNFYNGG---SPOGEAPGVQVFAAGAAAGVAATVSOCPQINLK-----NDS 266  
DB 406 GQCQACNDGFKNGDGAQCPHE-----SKTCSAGTA---SDCTECPTRALRYGDDG 455  
QY 267 P-----ATAGAQLATQCS-----TQCTPTGTAI-QDGV-----TLVFSNS 301  
DB 456 TKGTGCGECTGTGAGACKTCGLTIDGASYCSECATTTTEYPQNGVCAPKASRATPTCND 515  
QY 302 STQ---CSQCTANTYF-FNGN-FEA-----GKSQCLKCPVSKTTPAHAPGNATATQCTLT 352  
DB 516 PTQNGVCGTCADNPFKNGGCEYFKVPKTVCLISAPNGCTCOAKADGKLDSTG--LIV 573  
QY 353 CPAGTVLDGTSTNFVASATECTKCSAGFF--ASKTTGFTAGTDTCTECTKLTSGATAK 410  
DB 574 CSEG-----CRECASSTDCCTCLDGYVKSASACTKCDASCTCNGAATTCACATG- 624  
QY 411 VYAEATQKVQCAAS 423  
DB 625 YYKTASGEGACTS 637

RESULT 3  
T42017  
cysteine rich protein - Giardia intestinalis  
C:Species: Giardia intestinalis  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
C:Accession: T42017  
R:Chen, N.; Upcroft, P.; Upcroft, J.  
Parasitology 111, 423-431, 1995

Db 56 ASSNTKCKASADEASDQTCGKCLSTTFMYKGGCYDKTGNLRIICKTEGSDAGKGCACN 115  
 Qy 113 DGVVDVDFRSDAA-----QCVKCKPNFYNGGSPQGEAPGVQVFAAGAAA-AGVAAVTSQ 166  
 Db 116 DE--KGFFDNDPAANNVDSIC-----GDATGVTIPGSSTTKYKGVAGCAK 161  
 Qy 167 CVPQCLNKNDSPPATA---GAQANL-----ATQC--SNOCPTGTVDLDGVTLVFNT 211  
 Db 162 TKPSQISENTGTKEATCTECNANLYLKAVSSPTSAISVSAEDCKTG-----YFPT 212  
 Qy 212 SATL-----CVKCRPNFYNGGSPQGEAPGVQVFAAGAAAAGVAAVTSQVQCQINKNDS 266  
 Db 213 TDTTDSKKKCLTC-----STADKGGIDCSACELLPTSTRATVLIISCSACSIN-NLS 264  
 Qy 267 PATAGAQANLATOCSTOCPTGTALQDGVTLVFNSTQCSQCIANYFFNGNFEAKSQCL 326  
 Db 265 P-----LNKNECMQDCPAGTVADSNVC---KPCHTSCASCKGD-----NTE---SSCT 305  
 Qy 327 KCPVSKTTPAHAGTNTATQATQ-----CLTTCPCAGTVLD---DGTSTNFFVASATECTKCS 378  
 Db 306 AC-----YPGSVLSYGTDTNKGTCIAEC-TGKYLENCADGQCTATAGSKYCKCK 355  
 Qy 379 AGFF-----ASKTYGFTAGTDTCTCTCK--LTSGA----- 407  
 Db 356 SGFVPVNGLCVSAETARAAPPSTPDKTNG-----VCTACTEKYFLEGGCYQAEKPPG 409  
 Qy 408 -TARVYAEATQKVOCAS 423  
 Db 410 NTLCTTADAGKCTTCAN 426

RESULT 5  
 A45664  
 variant-specific surface protein vsp1267 - Giardia lamblia  
 C:Species: Giardia lamblia  
 C>Date: 22-Apr-1993 #sequence\_revision 18-Nov-1994 #text\_change 29-Jan-1999  
 C:Accession: A45664  
 R:Mowatt, M.R.; Aggarwal, A.; Nash, T.E.  
 Mol. Biochem. Parasitol. 49, 215-227, 1991  
 A:Title: Carboxy-terminal sequence conservation among variant-specific surface proteins  
 A:Reference number: A45664; MUID:92131058; PMID:1775165  
 A:Accession: A45664  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-596 <MOW>  
 A:Cross-references: GB:M63966; NID:g159140; PID:g159141  
 A:Note: sequence extracted from NCBI backbone (NCBIN:77609, NCBI:77610)

Query Match 9.0%; Score 210.5; DB 2; Length 596;  
 Best Local Similarity 22.6%; Pred. No. 2.9e-06;  
 Matches 125; Conservative 50; Mismatches 188; Indels 189; Gaps 33;  
 Qy 5 ILLIILISLFTNELRAVPCPD-GFQTQAGLTD-VGRADL-----GTC----- 44  
 Db 4 IAFYLILSTF-----AVDCKNSGNSCEAGQCDTIGDTEICMOCNQKVPINGICTAHSEE 58  
 Qy 45 -----VNCRPN-----FYNGGAAQ-GEANGN---QFFAANNAAR-- 75  
 Db 59 AVTNAGCKKNGTNIIESDKVQCQNGNYFLHGGCYKIGEPGNLICADEASNPCARTA 118  
 Qy 76 GICVPCQINRVGSVTNAGDILATQC-----STQCPT-GTALDDGVTDFD-----RS 123  
 Db 119 GVCAC---RDGYKNSDAVAT-ADSCIACEDANCATCGGAKENCTKICIDYFVGATGN 174  
 Qy 124 AAQCCKCK---PNFYNGG-----SPQGEAPGVQVFAAGAAAAGVAAVTSQCVP---C 170  
 Db 175 EGGCIKCDATGPNSTKGVAGCAKCBKPNAGPAKIECAADYLYKTEADEQTSVSEAVC 234  
 Qy 171 QLNKNDSPAT--AGAANLATOCNQCPGTGTVLDDGV-----TLVFNSTAT-----LCVK 218  
 Db 235 REGKTHFPTTDSAGGNKKVCVSC-----GTTNNGGIENCGECTSKESAARAGTEITCK 288  
 Qy 219 CRPNFYNGGSPQGEA-----PGVQVFAAGAAAAGVAAVTSQVQCQINKNDSPPATA--- 270

Db 289 CSS-----NNLSPLDGLACLTCPAGTVAVSGDSGVCKPCPCHTSCAGCQTDRETSCCTACYP 344  
 Qy 271 -----GAQANLATQCS-----TOCPTGTAIODGV----- 294  
 Db 345 GYSLLYESNGATGRVKECTGAFTTCNADGQCTANYGGAKYCTQCKDGVAPIDGICTAVA 404  
 Qy 295 -----TLVFNSTSTQCSQCIANY-----FFNGNFRAGKSQCL-----KCPVSKT--T 334  
 Db 405 AAGHDVSVCTATGKCTACTGNVALLSGCYNTQTLPGRKSVCKAVANSNDGKC---KTCA 461  
 Qy 335 PAHAPGNATQAQCLTTCPCAGTVLDGTTSTNFFVASATECTKCSAGFFASKTTGTGTAGTD 394  
 Db 462 NQAPDPATNFCPLCDSTCAECSTKND-----ADACTKCFPGYY---KTG-----N 504  
 Qy 395 TCTECTKKLTSG 406  
 Db 505 KCIKCTESSNNG 516

RESULT 6  
 A42125  
 trophozoite cysteine-rich surface antigen 170 - Giardia lamblia  
 N:Alternate names: CRP170; cysteine-rich surface antigen CRP170  
 C:Species: Giardia lamblia  
 C>Date: 05-Dec-1998 #sequence\_revision 05-Dec-1998 #text\_change 01-Dec-2000  
 C:Accession: A42125; B42125; S00530; S48056  
 R:Adam, R.D.; Yang, Y.M.; Nash, T.E.  
 Mol. Cell. Biol. 12, 1194-1201, 1992  
 A:Title: The cysteine-rich protein gene family of Giardia lamblia: loss of the CRP170  
 A:Reference number: A42125; MUID:92186850; PMID:1545800  
 A:Accession: A42125  
 A:Molecule type: DNA  
 A:Residues: 1-98 <ADAL>  
 A:Cross-references: GB:M83937; NID:g159124  
 A:Experimental source: trophozoite  
 A:Note: sequence extracted from NCBI backbone (NCBIN:88421, NCBI:88427); this ORF is  
 A:Note: the authors report but do not show 19 tandem repeats of the sequence of resid  
 A:Accession: B42125  
 A:Molecule type: DNA  
 A:Residues: 1269-1766 <ADA2>  
 A:Cross-references: GB:M83933; NID:g159122  
 A:Note: sequence extracted from NCBI backbone (NCBIN:88424, NCBI:88431); this ORF is  
 R:Adam, R.D.; Aggarwal, A.; Lal, A.A.; de la Cruz, V.F.; McCutchan, T.; Nash, T.E.  
 J. Exp. Med. 167, 109-118, 1988  
 A:Title: Antigenic variation of a cysteine-rich protein in Giardia lamblia.  
 A:Reference number: S00530; MUID:88089405; PMID:3335828  
 A:Accession: S00530  
 A:Molecule type: DNA  
 A:Residues: 1154-1409, 'A', 1411-1420, 'K', 1422-1425, 'R', 1427-1481 <ADA3>  
 A:Cross-references: EMBL:X06741; NID:g9355; PID:g929603  
 R:Yang, Y.; Adam, R.D.  
 Nucleic Acids Res. 22, 2102-2108, 1994  
 A:Title: Allele-specific expression of a variant-specific surface protein (VSP) of G1  
 A:Reference number: S48056; MUID:94301794; PMID:8029018  
 A:Accession: S48056  
 A:Molecule type: DNA  
 A:Residues: 1-56 <YAN>  
 A:Cross-references: EMBL:L25059  
 A:Experimental source: trophozoites WBA6  
 A:Note: the source is designated as Giardia intestinalis  
 C:Comment: This translation was produced by PIR staff from information provided by th  
 C:Genetics:  
 A:Gene: VSPA6  
 C:Keywords: surface antigen; tandem repeat

Query Match 8.8%; Score 206; DB 2; Length 1766;  
 Best Local Similarity 24.0%; Pred. No. 1.4e-05;  
 Matches 124; Conservative 44; Mismatches 185; Indels 164; Gaps 31;  
 Qy 20 AVPCPDGTQTAG-LTDVGAADLTGTCNCRPNFYNGGAA-----QGEANGNPFAAN--- 71  
 Db 1101 AVDC-----QSGAGYYTDDSDSAKECKKCNAPCTACGTADKCTCDANGAAPYLKKTNP 1156

1027 MTVPNDST-TTAPSSSTVNSASSETPCNSETQTSDDGTSTMTVSSDSTTPGSPSTATNSAS 1085  
 QY 219 CRPNFYNGSGPQGGAPGVQVFAAGAAAAGAAVTSQVPCQIN-----KNDS 266  
 Db 1086 CETPFVYSHKFKYDST-----TAGPSSSTAINSSASE-TPCNSETQTSDDGTSTMTVPND 1139  
 QY 267 PATAGAANLATQCS--TQCPTGTATQDGV-TLVFNSSTOCSQCIANYFFNGFNFEAGS 323  
 Db 1140 -TTAGPSSSTVNSASSETPCNSETQTSDDGTSTMTVSSDST-----TAGPS 1183  
 QY 324 QCLKCPVSKTTPAHA-----PNTATQATQCLTTPAGTVLDDG 362  
 Db 1184 STVNSASSETPCNSETQTSDDGTSTMTVSSDSTTAGPSSSTVNSASSETPCNSETQTSDDG 1243  
 QY 363 TSTNFV-----ASATECTKCSAGFFASKTTGTAGTGTCTECCKLTSGATAKYVAE 414  
 Db 1244 TSTMTVPNDSTTAGPSSSTVNSASSETPCNSETQTSDDGTSTMTVSSDSTTAGPSSSTVNS 1303  
 QY 415 ATOKVQCASTT 425  
 Db 1304 ASSETPCNSET 1314

RESULT 8  
 C42125  
 trophozoite cysteine-rich surface antigen 72 - Giardia lamblia (fragment)  
 N:Alternate names: CRP72  
 C:Species: Giardia lamblia  
 C:Date: 05-Dec-1998 #sequence\_revision 05-Dec-1998 #text\_change 05-Dec-1998  
 C:Accession: C42125  
 R:Adam, R.D.; Yang, Y.M.; Nash, T.E.  
 Mol. Cell. Biol. 12, 1194-1201, 1992  
 A:Title: The cysteine-rich protein gene family of Giardia lamblia: loss of the CRP170  
 A:Reference number: A42125; MUID:92186850; PMID:1545800  
 A:Accession: C42125  
 A:Molecule type: DNA  
 A:Residues: 1-677 <ADA>  
 A:Cross-references: GB:M83934; NID:gi159123  
 A:Experimental source: trophozoites  
 A:Note: sequence extracted from NCBI backbone (NCBIN:88443, NCBI:88444); this ORF is  
 C:Keywords: surface antigen

Query Match 8.7%; Score 203; DB 2: Length 677;  
 Best Local Similarity 24.6%; Pred. No. 9.6e-06;  
 Matches 101; Conservative 29; Mismatches 179; Indels 102; Gaps 21;

QY 31 AGLTDVGAADL-----GTCVNCRPNFYNGGAAQGEANGNQPFAANNAARGICVPCQINRV 86  
 Db 54 AGCTKAGGAALDKMTATCEKCDGDFLFWGCKYKTTDGPGESEICTKAEGLCTEKTAN- 112  
 QY 87 GSVTNAGDLATLATQCTQCTPTGTALDDGVTDVDFRSDAAQCVKCKPNFYNGGSPQGEAP 146  
 Db 113 GLFKNPAATPEKGSBCLCSDINGA--DGYTGV--ANCAOCTKSDSN----- 155  
 QY 147 GVOVFAAGAAAAGAAVTSQVPCQINKNDSPATAGAANLATQCSNOCPTGTVLDDGVT 206  
 Db 156 -----KGAAT-----CTACQ-----AGYKDFP--QACSK-----DGTG 182  
 QY 207 LVFNTSATLVCVCRPNFYNGGSPQGEAPGVQVFAAGAAAAGAAVTSQVPCQINKNDS 266  
 Db 183 LTCETSAOCTSCPEKYLKDKSCVNNNG-----CTGNTYAD--PESGKCLPC--NTIDQ 234  
 QY 267 PAT-----AGAANLATQCTQCTPTGTATODGVTLVFSNSTQCSQCIANYFFNGFNFEAG 321  
 Db 235 ACTQCEVDSTTKPKCTKCGGKWKWTAI-DGTTTCVDANGCATSNVDSHFLND-----G 289  
 QY 322 KSQCLKC-PVSKTTPAH--APGNTATQATQCLTTCP-----AGTVLDDGTSTNF 367  
 Db 290 STKILCSDSSLEANKGTPGCKTKKNGAKPTCECLDGYNSGNGGTVTCCEAGANC 349  
 QY 368 V----ASATECTKCSAGFFASKTTGFT-----AGTDTCTECKLT 404  
 Db 350 ATCTQAGNDKCTCKPKGFF--MKGNGPTGECVACDQAQGGIDGCAECTREST 399

72 NAARGIC--VPCQINRVG-----SVTNAGDLATLATQCS-----TQC----- 106  
 Db 1157 SDPTGTCVSAVDQ--GSAGYYTDDSVSDAKCKNACPTACAGTADKCTKDANGAAPY 1215  
 QY 107 -----PTGTAL-----DDGVTDVDFRSDAAQCVKCKPNFYNGGSPQGEA 145  
 Db 1216 LKKNPSDPTGTCVSAVDQGSAGYYTDDSVSD-----AKECKN----- 1256  
 QY 146 PGVOVFAAGAAAAGAAVTSQVPCQIN-----KNDSPATAGAANLATQCSNOCPTGT 199  
 Db 1257 -----APCTACAGTA-----DKCTKDANGAAPYLKTNPSDPTGTCVSNVDC--QGSAGY 1305  
 QY 200 VLDDGV-----TLVFNSTATLVCVCRPNFY-----YNGGSP 230  
 Db 1306 YTDSSVDAKECKKCBGQKPNATAGTQCFSCDANCERCQDNDVCARCSTGAPPENKCP 1365  
 QY 231 QGEAPGVQVFAAGAAAAGAAVTSQVPCQINKNDSPATAGAAN--LATQCSQCTPTG 287  
 Db 1366 -AATPGCHSSCDGCTENAMTNQADKCTGCKEGRYLKPEAAGSGTCLTAECTSDTHF 1424  
 QY 288 T-----ATQDGVTLVFNSS--TQCSQCIANYFFNGFNFEAGKSQCLKCPVSKTTPAHPG 340  
 Db 1425 TKEAGDSKGNCLSCSDATHGIGCKK--ALKTLGSAES--TVVCSCTDKRLTPS--G 1479  
 QY 341 NTATQATQCLTCTPAGTVLD--DGTST-----NFVASATCTKCSAGFFASKTT 387  
 Db 1480 NA-----CLDNCBPAGTYADNINGSVCASCHATCAECNDANAASCTACYPGYSILYGS 1533  
 QY 388 GFTAGTDTCT--ECTKK--LTSGATAKYVAEATQKVOCA 422  
 Db 1534 G-TAG--TCVTECTGAFITNCADGCTANVGGAKYCA 1567

RESULT 7  
 T32271  
 hypothetical protein ZC178.2 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 C:Accession: T32271  
 R:Murray, J.; Wohldmann, P.; Beck, C.  
 submitted to the EMBL Data Library, September 1997  
 A:Description: The sequence of C. elegans cosmid ZC178.  
 A:Reference number: Z21143  
 A:Accession: T32271  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1459 <MUR>  
 A:Cross-references: EMBL:AF024496; PIDN:AAB70340.1; GSPDB:GN00023; CESP:ZC178.2  
 A:Experimental source: strain Bristol N2; clone ZC178  
 C:Genetics:  
 A:Gene: CESP:ZC178.2  
 A:Map position: 5  
 A:Introns: 1099/2

Query Match 8.8%; Score 205; DB 2: Length 1459;  
 Best Local Similarity 21.6%; Pred. No. 1.4e-05;  
 Matches 106; Conservative 56; Mismatches 205; Indels 124; Gaps 18;

QY 22 PCPDGTTQAGL-----TDVGAADLGTVCNCRPNFYNGGAAQGEANGNQPFAANN 72  
 Db 861 PCNSETQTSDDGTSTMTVPNDSTTAGPSSSTAT-----NSASSETPCNSETQTSDD 908  
 QY 73 AARGICVPCQINRVGVTNAGDLATLATQCS--TQCPTGT-----ALDDGVTDVDFR 122  
 Db 909 GTSTMTVPNDSTTAGPSSSTATNSASSETPCNSETQTSDDGTSTMTVPNDSTTAGPSSSTATN 968  
 QY 123 SAAQCVKCKPNFYNGGSPQGEAPGVQVFAAGAAAAGAAVTSQVPCQIN----- 173  
 Db 969 SASSETPCNSETQTSDDGTSTMTVPNDST-TAGPSSSTATNSASE-TPCNSETQTSDDGTST 1026  
 QY 174 ---KNDSPATAGAANLATQCSNOC-----TGTVLDDGVTLVFNSTATLVCVK 218

## RESULT 9

Tl0053  
laminin alpha 5 chain - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 11-Jan-2000  
C:Accession: Tl0053  
R:Miner, J.H.; Lewis, R.M.; Sanes, J.R.  
submitted to the EMBL data Library, November 1997  
A:Reference number: Z16923  
A:Accession: Tl0053  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-3635 <MIN>  
A:Cross-references: EMBL:U37501; NID:g2599231; PID:g2599232  
C:Genetics:  
A:Gene: Lama5  
C:Superfamily: unassigned EGF-related proteins; EGF homology; laminin-type EGF-like homology  
C:Keywords: basement membrane; cell binding; extracellular matrix  
F:1888-1939/Domain: laminin-type EGF-like homology <LEG>  
F:1942-1970/Domain: EGF homology <EGF>

Query Match 8.6%; Score 202.5; DB 2; Length 3635;  
Best Local Similarity 20.0%; Pred. No. 4.4e-05;  
Matches 94; Conservative 36; Mismatches 150; Indels 191; Gaps 21;

QY 43 TCVNCRPNEY-----YNGGAAGGANGNQPFAANNAARGTCVPCQINRVGSVTN----- 91  
Db 1759 SQCECAPGYRTKGLFLGRVPCQCHGHSRCLPGS--GTCVGCQHNTGDCQERCRCGP 1816  
QY 92 --AGDLATLATQC-STQCP-----TGTALDDGVTDVDFR---SAAQCVKCKPNFY 136  
Db 1817 FVSSDPSPNPASPCVSCPCPLAVPSNNFADGCVLRNGRTQCLRCRGCYAGASCERCAPGFF- 1875  
QY 137 NGSGPQGEAPGVQVFAAGAAAGVAATVSCQVPCQLNKNKNSPATAQAANLATQCSNQCP 196  
Db 1876 --GNP-----LVLGSSCPDCSGNGDPNMFISDCDPLTGACRGC- 1913  
QY 197 TGTVLDDGVTLVFNSTFLVKCRPNFYNGSGPQGEAPGVQVFAAGAAAGVAATVSCQ 256  
Db 1914 -----LRHTTGPCHERCAPGV-----GNALLPGNC 1939  
QY 257 VPCQINKNSPATAGAAANLATQCSQTGTAIQD-----GVTLVFNSSTQCS 306  
Db 1940 TRC-----DCS-----PCGTETCDPQSGRCLCKAGVT-----GQRCD 1971  
QY 307 QCIAFYFNGNFAGKSOCLKC-----PVSKTTPAH-----APGNTATQATQCL-- 350  
Db 1972 RCLGEGYF-----GFQCCQCRPCACGPAKGSCHQPSQCHQCPGTGTPCQLECAPG 2024  
QY 351 -----TTCPAGTVLDDGTSTNFNVASATECTKSAGF---F 382  
Db 2025 YWGLPEKGRRCQCPRGCHDHPHTGHCTCPGP-----LSGERCDTCSQQHQPVP 2072  
QY 383 ASKTTGTAGTDTCTECTKKL-----TSGATAKVAEATQKYOCATTTAK 428  
Db 2073 PGKPGGHGHCVECHVCLLDLRLERAGALLPAIREQLQGINASSAAWAR 2123

## RESULT 10

MMMSA  
laminin alpha-1 chain precursor - mouse  
N:Alternate names: laminin chain A1  
C:Species: Mus musculus (house mouse)  
C>Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 19-Jan-2001  
R:Sasaki, M.; Kleiman, H.K.; Huber, H.; Deutzmann, R.; Yamada, Y.  
J. Biol. Chem. 263, 16536-16544, 1988  
A:Title: Laminin, a multidomain protein. The A chain has a unique globular domain and h  
A:Reference number: A31771; MUID:89034134; PMID:3182802  
A:Accession: A31771  
A:Molecule type: mRNA

A:Residues: 1-3084 <SAS>  
A:Cross-references: EMBL:J04064; NID:g309419; PIDN:AAA39410.1; PID:g309420  
A:Accession: A30449  
A:Molecule type: protein  
A:Residues: 183-195;570-571,'A',573-586;596-612,'X',614-617,'EMK',630-646;1217-1222,'  
2486;2624-2639;2818-2843;3009-3033,'V',3035 <SA2>  
R:Hartl, L.; Oberbaeumer, I.; Deutzmann, R.  
Eur. J. Biochem. 173, 629-635, 1988  
A:Title: The N terminus of laminin A chain is homologous to the B chains.  
A:Reference number: S00624; MUID:88225080; PMID:3267223  
A:Accession: S00624  
A:Molecule type: mRNA  
A:Residues: 1-208,'T',210-334 <HAR>  
A:Cross-references: EMBL:X07737; NID:g52857; PIDN:CAA30561.1; PID:g52858  
A:Accession: A30450  
A:Molecule type: protein  
A:Residues: 311-335,'N',337-339;630-642,'D',644-692-734;737-748,'X',750-760,'G',762-7  
3-1389;1449-1459 <HA2>  
A:Note: the sequence from Fig. 7 is inconsistent with that from Fig. 5 in having 209-  
R:Mann, K.; Deutzmann, R.; Timpl, R.  
Eur. J. Biochem. 178, 71-80, 1988  
A:Title: Characterization of proteolytic fragments of the laminin-nidogen complex and  
A:Reference number: S08895; MUID:89078415; PMID:2462498  
A:Accession: S08895  
A:Molecule type: protein  
A:Residues: 153-169 <MAN>  
R:Fujiwara, S.; Shinkai, H.; Deutzmann, R.; Paulsson, M.; Timpl, R.  
Biochem. J. 252, 453-461, 1988  
A:Title: Structure and distribution of N-linked oligosaccharide chains on various dom  
A:Reference number: S02678; MUID:88326259; PMID:2458101  
A:Accession: S02678  
A:Molecule type: protein  
A:Residues: 630-642,'D',644-2690-2704 <FUJ>  
R:Deutzmann, R.; Huber, J.; Schmetz, K.A.; Oberbaeumer, I.; Hartl, L.  
Eur. J. Biochem. 177, 35-45, 1988  
A:Title: Structural study of long arm fragments of laminin. Evidence for repetitive C  
A:Reference number: S01790; MUID:89030693; PMID:3181157  
A:Accession: S01790  
A:Molecule type: mRNA  
A:Residues: 2538-3084 <DEU>  
A:Cross-references: EMBL:X13459; NID:g55499; PIDN:CAA31807.1; PID:g818014  
A:Accession: A30451  
A:Molecule type: protein  
A:Residues: 1911-1929;1997-2006;2033-2045,'X',2047-2054,'X',2056-2066,'X',2068-2105;2  
470-2487-2498;2502-2525;2538-2557;2561-2591,'X',2593-2594;2600-2610;2616-2645;2648-26  
93;2998-3005,'A',3007-3033,'V',3035;3068-3083 <DE2>  
A:Note: 2256-Val was also found  
R:Olson, D.; Nagayoshi, T.; Fazio, M.; Peltonen, J.; Jaakkola, S.; Sanborn, D.; Sasak  
Lab. Invest. 60, 772-782, 1989  
A:Title: Human laminin: cloning and sequence analysis of cDNAs encoding A, B1 and B2  
A:Reference number: A34961; MUID:89280632; PMID:2733383  
A:Accession: S14670  
A:Molecule type: protein  
A:Residues: 2424-2436;2440-2451;2461-2467;2487-2525;2550-2557;2561-2593;2600-2610;261  
-2942,'T',2944-2964;2969-2976;2980-2993;2998-3000,'I',3002-3018,'V',3020-3034;3068-30  
C:Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type lamin  
C:Function:  
F:25-277/Domain: laminin alpha-1 chain #status predicted <MAT>  
F:25-277/Domain: VI <DOM6>  
F:277-331/Domain: laminin-type EGF-like homology <LE01>  
F:278-519/Domain: V <DOM5>  
F:334-401/Domain: laminin-type EGF-like homology <LE02>  
F:404-458/Domain: laminin-type EGF-like homology <LE03>  
F:461-507/Domain: laminin-type EGF-like homology <LE04>  
F:510-519/Domain: laminin-type EGF-like homology <LE05>  
F:520-715/Domain: laminin-type EGF-like homology #status atypical <LE05>  
F:716-1166/Domain: IIB <DOM4>  
F:716-746/Domain: laminin-type EGF-like homology #status atypical <LE06>  
F:749-795/Domain: laminin-type EGF-like homology <LE07>

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C/Accession: T25933

R/Murray, J.; Wohldmann, P.

submitted to the EMBL Data Library, December 1996

A/Description: The sequence of C. elegans cosmid W02C12.

A/Reference number: T20112

A/Accession: T25933

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1372 <MUR>

A/Cross-references: EMBL:U80815; PIDN:AAB37995.1; GSPDB:GN00022; CESP:W02C12.1

A/Experimental source: strain Bristol N2; clone W02C12

C/Genetics:

A/Gene: CESP:W02C12.1

A/Map position: 4

A/Introns: 29/1; 66/1; 774/2; 823/2; 1046/1; 1108/2; 1298/1

Query Match 8.3%; Score 195.5; DB 2; Length 1372;

Best Local Similarity 23.9%; Pred. No. 5.3e-05;

Matches 105; Conservative 43; Mismatches 155; Indels 137; Gaps 27;

30 QAGLTDVGAADLGTVCNCRPNFYNGGAAQGEA-----NGNQFAANNAARGICVPCQIN 84

Db 923 QAG--QVVRDL--CVPCAPCTYHS--AATGECELCPIGEYQPL-----TARTECFKC--- 969

QY 85 RVGSVTNAGDLATLQCSQCTGTTALDDGVTDFDRSAAQCVCCKPNFYNGGSPQGE 144

Db 970 APQGIT-ASEGAISEGECKDNCPFGHQYDLSLTS-----CVTCGYGY-----QP--- 1013

QY 145 APGVQVFAAGAAAGAAVTSQCVPCOLKNDSPATAGAAQANLATQCSNOCPTGTVLDDG 204

Db 1014 -----SAGA-----FECIPCGIGK-----TTLSEFATSEDCRDECPDGEQL--- 1050

QY 205 VTLVFNATLVCVKCRPNFYNGGSPQGEAPGVQVFAAGAAAGAAVTSQCVPCQINKN 264

Db 1051 -----SASGVCCOPCOIGTYRSRG-----NKKVCAC----- 1076

QY 265 DSPATAGAAQANLAT---OCST--OCPTGTAIQDGTVLVFNSTQCSOCIANYFENGFEA 320

Db 1077 --PPGTTTATMSTRREQNTPKCPQGFL-----VKETKNCQFCPRGTQFN---EE 1123

QY 321 GKSOCLKCPVSTTPPAHAPGNATQATQCLTT--CPAGTV-----LDDGTSTNFVASA 371

Db 1124 QESTCKLCPADHTTA--APG--ATAESQCFSTNQCATGEYNCNWHANCIDLDPDENDVPS- 1178

QY 372 TECTKSAGFFASKITGFTAGTDTG-----TEC--TKKLTSGATAKYVAEATQ 417

Db 1179 YEC-RCKPGYRGNTHCTDACNDLNDGICKKNNIGNVECLCKDHFSDRCLELRFQASN 1237

QY 418 KVOCASITTFAKFLSISLFI 437

Db 1238 NKLWIAIVGVVVIGIIV 1257

RESULT 12

T42215

zonadhesin - mouse

N/Alternate names: sperm-specific membrane protein

C/Species: Mus musculus (house mouse)

C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

C/Accession: T42215

R/Gao, Z.; Garbers, D.L.

J. Biol. Chem. 273, 3415-3421, 1998

A/Title: Species diversity in the structure of zonadhesin, a sperm-specific membrane

A/Reference number: 22080; MUID:98123114; PMID:9452463

F:798-853/Domain: laminin-type EGF-like homology <LE08>

F:830-834/Region: cell adhesion #status predicted

F:856-906/Domain: laminin-type EGF-like homology <LE09>

F:909-955/Domain: laminin-type EGF-like homology <LE10>

F:958-1002/Domain: laminin-type EGF-like homology <LE11>

F:1005-1048/Domain: laminin-type EGF-like homology <LE12>

F:1051-1094/Domain: laminin-type EGF-like homology <LE13>

F:1097-1116/Domain: laminin-type EGF-like homology #status atypical <LE14>

F:1118-1154/Domain: laminin-type EGF-like homology #status atypical <LE15>

F:1147-1149/Region: cell attachment (R-G-D) motif

F:1157-1166/Domain: laminin-type EGF-like homology #status atypical <LE16>

F:1167-1368/Domain: IVA <D04A>

F:1369-1561/Domain: IIIa <D03A>

F:1389-1407/Domain: laminin-type EGF-like homology #status atypical <LE17>

F:1410-1456/Domain: laminin-type EGF-like homology <LE18>

F:1459-1513/Domain: laminin-type EGF-like homology <LE19>

F:1516-1560/Domain: laminin-type EGF-like homology <LE20>

F:1562-2133/Domain: II/I <DOM2>

F:1562-2133/Region: heptad repeats

F:2134-3084/Domain: G <DOM3>

F:2150-2308/Domain: laminin G repeat homology <LG1>

F:2337-2492/Domain: laminin G repeat homology <LG2>

F:2518-2683/Domain: laminin G repeat homology <LG3>

F:2748-2897/Domain: laminin G repeat homology <LG4>

F:2925-3082/Domain: laminin G repeat homology <LG5>

F:25/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted

F:45-79,370,374,531,562,672,808,914,959,969,1052,1344,1414,1586,1603,1659,1686,1706,1718

F:45/ (covalent) #status predicted

F:304-312/Disulfide bonds: #status experimental

F:770-857,1999,2055,2067,2835/Binding site: carbohydurate (Asn) (covalent) #status experi

F:845,2102/Binding site: carbohydurate (Asn) (covalent) #status absent

Query Match 8.5%; Score 198; DB 1; Length 3084;

Best Local Similarity 20.0%; Pred. No. 7.3e-05;

Matches 89; Conservative 38; Mismatches 142; Indels 176; Gaps 16;

QY 23 CPDG-TQTQAGLTDVGAADLGTVCNCRPNFYNGGAAQGEANGNOPFAANNAARGICVPC 81

Db 718 CPQGYGT-----SCEACLPGYRVGILFG-----GICQPC 749

QY 82 QINRVGSVTNAGDLATLQCS-----OCPTG----- 109

Db 750 ECHGASECDIHGICSVCTHTTGDHCEQLPGFYGTSPRGTPGDCPCACPLSIDSNFN 809

QY 110 ----TALDDGVTDFDR-----SAAQCVCCKPNFYNGGSPQGEAPGVQVFAAGAAAGVA 161

Db 810 SPTCHLTGDEVVYCDQCAPGYSGSWCERCADGYGNPTVPG----- 851

QY 162 AVTSQCVPCOLKNDSPATAGAAQANLATQCSNOCPTGTVLDDGVTLVFNATLVCVKCRP 221

Db 852 ----TCVPCNCSGNVDLEAGHCDSTGEC-----LKCLWNTDGAHCERCAD 894

QY 222 NFYNGGSPQGEAPGVQVFAAGAAAGAAVTSQCVPCQINKNDSPATAGAAQANLATQCS 281

Db 895 GFY-----GDAVTAKNCRACDCHENG-----LS 918

QY 282 TQCTGTAIQDGTVLVFNSTQCSOCIANYFENGFEAGSKOCLKCPVSKTTPPAHAPGN 341

Db 919 GVCHLEGLDCKRPHV---TGQQCDQLSGY-----GLDTGLGCVPCNCSVEGVSVD 968

QY 342 TATQATQCLTTCPTAGTVDLDDGTSTNFVASATECTKSAGFFASKITGFTAGTDTCTCTK 401

Db 969 NCTEEGQC--HCGPG-----VSGKQCDRCRSHGFYAFQDGG-----CUPDC 1007

QY 402 KLTSGATAKYVAEA-----TKQVQC 421

Db 1008 AHTQNNCDPASGECLCPPTHTQGLKC 1032

RESULT 11

T25933

hypothetical protein W02C12.1 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans





Search completed: February 11, 2003, 19:48:39  
Job time : 28.0286 secs

```
Db 881 PPAYADKKRKCMPVG-----CSTCTSAFLSCPEKWLN---KKGKMPV-- 926
QY 240 FAAGAAAGVAAVTSQVPCOINKNDSPATAGAAANLATOCSTCPTGTATIODGVTLVFS 299
Db 927 -GSDKCSAGFAVDKCKRC-----NPACDSCYGENEGHCLT-CNPNNLQD----- 971
QY 300 NSTQC-SOCIANTYFFNGFAGKSOCLKCP-----VSKTTPAHAGNTATQATCLTT 352
Db 972 ---YKCVPECSKGYI-----AEAGR--CARCMHGSCDCVSRINCTSCASTLRLQSGACRTS 1022
QY 353 CPAGTVLDDGTSTNFVASATECKCSAGFFASKTTGTTAGTDICTECTK--KLTSG 406
Db 1023 CADGYADRGTT-----CSKC-----YLSCRTICGPRDQCASCPEGWRLAAG 1064

RESULT 15
A48579
trophozoite surface protein TSP11 - Giardia lamblia
C:Species: Giardia lamblia
C>Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 29-Jan-1999
C:Accession: A48579
R:Ey, P.L.; Khanna, K.K.; Manning, P.A.; Mayrhofer, G.
Mol. Biochem. Parasitol. 58, 247-257, 1993
A:Title: A gene encoding a 69-kilodalton major surface protein of Giardia intestinalis
A:Reference number: A48579; MUID:93241215; PMID:8479449
A:Contents: Ad-1
A:Accession: A48579
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-667 <EYL>
A:Cross-references: GB:M95814; NID:g159106; PID:g159107
A>Note: sequence extracted from NCBI backbone (NCBIN:130056, NCBIPI:130058)

Query Match 7.8%; Score 182; DB 2; Length 667;
Best Local Similarity 22.2%; Pred. No. 0.0002;
Matches 120; Conservative 42; Mismatches 170; Indels 208; Gaps 32;

QY 7 LILILSLFELNLRVPCPDGTQT-----QAGLIDV-----GAADL-GTCVNCRPNFY 52
Db 1 MLLAIFYFVISTLAKTTQTCTCEAKCEMVGETEICRCQTKGKVPIDGKCVDTAN-- 58
QY 53 YNGGAOGEANGNOPFAANNAARG--ICVPCQINRVGSVT-----NAGDLATLATQC 102
Db 59 ANCNASGDDGANGQVCGKMSVPGNTLCTTVSPDGVCSVAANEYFVPPNADATHDSVWSC 118
QY 103 STQCPGTGALDDGTVDFDRSAOCVKCKPNFYNGSGPQGEAPGVQVFAAGAAAGVAA 162
Db 119 SEETPIHLANNKOYIGV-----AGCATC-----SAP--KAPGEDNTPKAA-- 157
QY 163 VTSQVPCQLNKNDSPATAGAAANLATQCSNOCPGTGTGLDDGVTLVFNFSAT-----LCV 217
Db 158 -----CTKCAAGFLHTPSEG-----LSSCEETCPEG-----YFGHTATAESKKTCK 198
QY 218 KCRPNFYNGSGPQGEAPGVQVFAAGAAAGVAAVTSQVPCOINKNDSPATAGAAANLA 277
Db 199 SC-----TGGG--SEAPNVK-----GI-----GDCLKCMYNE-----ASGNTL 229
QY 278 T--QCSTQ-----CPTGTAIODGVTLVFNFSNSTQ- 304
Db 230 TCEKCSAOKKPSLDKTSNCDTGNCAFCPSSSGGDCGCDSG-FILDGQNCVKSDCKTEN 288
QY 305 -----CSQCIANYFFN-----GNFEG-----KSQCLKCPVS--K 332
Db 289 CKACTNPRAANEVCTECISTHLLTPTSOCVOYCOALGNYYAGTNADNKKACECTVANCK 348
QY 333 T-----TPAHAPGNT--ATQATQCLTTCPCAGTVL---DDGTSTN 366
Db 349 TCNDQGCQCQTCNDGYKNGSDACSPCHESCKTCSAGTASDC-TECPTGKALKYGNQTK-- 405
QY 367 FVASAETCKCSAGFFASKTTGFTA-GTDTCTECTKLTSGATAKVAEATQKVQCASTT 425
Db 406 --GTCGEGCTTGCGGACKTCGLTIDGASYCSECDTQ-----NEYPNIGICTSTT 453
```